

Reliability of Precipitation Probabilities Estimated From the Gamma Distribution¹

T. C. BRIDGES and C. T. HAAN—Agricultural Engineering Department,
University of Kentucky, Lexington, Ky.

ABSTRACT—A technique is developed for evaluating the reliability of precipitation estimates determined by using the gamma distribution. Tables are presented showing the probabilities of errors of various magnitudes in precipitation estimates as a function of record length for selected cases.

1. INTRODUCTION

As the use of precipitation probabilities becomes more prevalent, the question of the reliability of such probabilities is increasingly important. The number of observations needed to reliably predict precipitation probabilities is an important consideration. Weisner (1970) indicates that from 25 to 50 observations of precipitation data are needed to give a "stable" frequency distribution.

Previous work in this area indicates that rainfall probabilities can be estimated from the gamma distribution. Thom (1958) developed methods of estimating the parameters of this distribution. Friedman and Janes (1957) made use of his methods in the estimation of rainfall probabilities for Connecticut. Other applications were made by Barger et al. (1959) and Strommen and Horsfield (1969) in predicting rainfall probabilities for different areas of the United States. Greenwood and Durand (1960) presented approximations for estimating the maximum likelihood parameters for the gamma distribution. Shenton and Bowman (1970) discuss some properties of Thom's estimators for the gamma distribution. In using the gamma distribution, these investigators made little mention of the reliability of the precipitation estimates. Friedman and Janes (1957) placed confidence intervals on the estimates of the parameters for the density function. Placing confidence intervals upon the parameters, however, does not fix the reliability of the precipitation estimates. Mooley and Crutcher (1968) investigated the number of years of record needed to stabilize the gamma parameters in a study of rainfall in India. This report sets forth a method for determining the reliability of the precipitation estimates as a function of the number of observations available for analysis.

2. PROCEDURE

Basically, the method consists of the following four steps:

1. Simulate p sets of rainfall data from the gamma distribution with known parameters. Each set of data consists of n observations.

2. Determine parameters of the gamma distribution that best fit the n observations for each set of data. This results in p sets of parameters. These parameters should equal the original parameters except for sampling fluctuation and any bias in the estimation procedure.

3. Determine the amount of rainfall that is expected to be exceeded $(100-x)$ percent of the time for each of the p parameter sets. This results in p estimates of rainfall at the x -percent probability level.

4. Determine the distribution of the p rainfall estimates at the x percent probability level and the reliability of a single estimate based on n observations.

This procedure, with the help of a computer program (Bridges and Haan 1971), was repeated for various values of the original known gamma parameters and for various values of n , the number of observations. A detailed explanation of each step follows.

Step 1

According to Hahn and Shapiro (1967), random values, y , that follow the gamma distribution can be simulated from

$$y = -\beta \sum_{i=1}^m \ln(1-R_u) \quad (1)$$

where R_u is a random number between 0 and 1 and β and m are the scale and shape parameters, respectively, of the gamma distribution. Initially, β and m are known and are read into the computer program. Using eq (1) and a computer-supplied subroutine, Randu (International Business Machines 1970), the program generates p sets of rainfall data each consisting of n observations. Randu, a random number generator that gives values between zero and one, is available in the IBM Scientific Subroutine Package.²

Step 2

For each of the p sets of simulated rainfall, the gamma parameters, β and m , are found. Thom (1958) gives procedures for estimation of these parameters from n observations of rainfall. The parameter m is found from the

¹ This research was supported by the Kentucky Agricultural Experiment Station and is published with the approval of the director as Paper No. 71-2-123.

² Mention of a commercial product does not constitute an endorsement.

quadratic relationship

$$12 \left(\ln \bar{y} - \frac{1}{n} \sum \ln y \right) m^2 - 6m - 1 = 0 \quad (2)$$

where \bar{y} is the arithmetic average of the n observations and $\sum \ln y$ is the sum of the natural logarithms of the same n observations. It should be noted that \bar{y} and $\sum \ln y$ are based on years when there was enough rain to record. This did not affect the estimation in this report, for only nonzero values are generated by eq (1). Equation (2) was solved by means of the quadratic formula, and only the positive roots were used. Once m is found, then β is given as

$$\beta = \frac{\bar{y}}{m} \quad (3)$$

Equations (2) and (3) are used to estimate the parameters of each of the p data sets generated in step 1.

Step 3

After estimating the p sets of parameter, one can then calculate probabilities for each set of parameter values using the incomplete gamma distribution. The gamma distribution is given by

$$p(y) = \frac{e^{-y/\beta} y^{m-1}}{\beta^m \Gamma(m)} \quad (4)$$

where y is the random variable (rainfall depth), β and m are the gamma parameters, and Γ is the gamma function. The probability of a given y value is zero, but the probability of exceeding y is one minus the area from zero to y under the distribution function. The area is obtained by numerically integrating eq (4) from zero to y . Values of y corresponding to probabilities of 95 and 99 percent were determined in this manner for each set of simulated data.

Step 4

A probability distribution is fit to the rainfall amounts at the 95- and 99-percent probability levels as determined in step 3. We found for most cases, through the chi-square test, that the log normal distribution adequately describes these data at the two probability levels for each of the original gamma parameters. Further investigation may indicate that some other distribution should be used for this step. The true rainfall amount, R , can be evaluated from eq (4) with the original, known parameters. If one knows the true rainfall amount and the standard deviation of the logarithms of the estimates of R based on the simulated data, a degree of reliability can be stated as to the probability of an estimate falling outside a specific limit or a specified deviation from the true rainfall amount for a given sample size.

If the data truly follow the gamma distribution, the shaded area of figure 1 represents the percent chance that an error of $\pm d$ exists in the estimate of the rainfall amount that is equaled or exceeded (100- x) percent of the time.

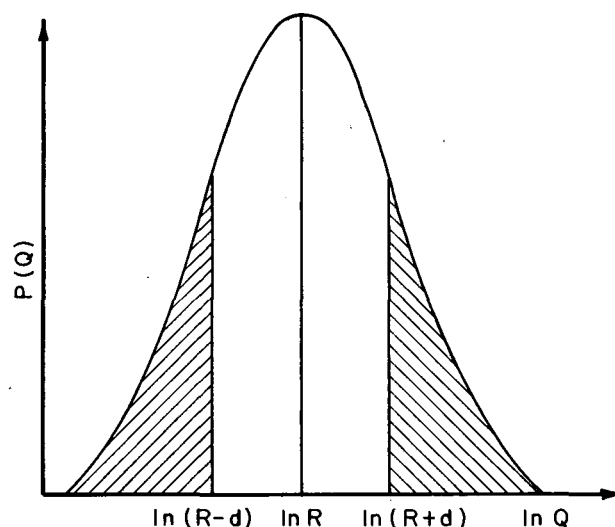


FIGURE 1.—Distribution of estimated rainfall amounts at a given probability level.

TABLE 1.—Population rainfall estimates, R

m	β	Probability level	
		95%	99%
1	1.00	3.10	5.29
2	1.69	8.02	11.21
3	0.26	1.64	2.19

In this report, p is 100 and n took on values of 10, 20, 30, 40, 50, and 100 observations. Three sets of β and m were used with four replications each, giving a total of 12 separate computer runs. Beta was set at values of 1.00, 1.69, and 0.26 while m took on values of 1, 2, and 3, respectively. These parameter values were selected arbitrarily; however, they are representative of values typically found in analyzing rainfall in the Eastern United States (Strommen and Horsfield 1969). Table 1 presents the population rainfall at the 95- and 99-percent probability levels for the three sets of gamma parameters used.

3. EVALUATION OF ERROR PROBABILITIES

Curves were prepared relating the number of observations to the standard deviation of logarithms of the estimated rainfall amounts for each probability level, x . Standard deviations were plotted for each replication, and a curve connecting the means of the replications was drawn. In some cases, the curve did not follow the mean standard deviation exactly because of a lack of sufficient replications to stabilize the mean standard deviation and present a smooth curve. Figure 2 is a typical example.

In all cases, these curves show a definite decrease in the standard deviation with an increase in the number of observations. The curves show a level of reliability in that values based on fewer observations are not as stable as evidenced by the larger standard deviations. Probability estimates based on fewer observations would not be as

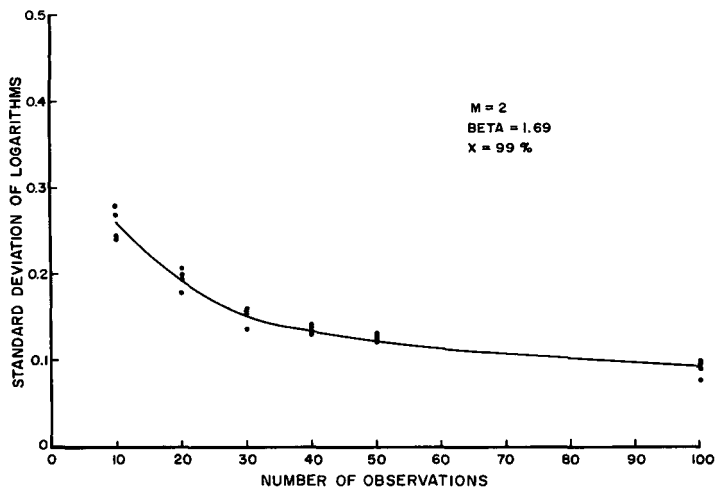


FIGURE 2.—Standard deviation of logarithms of precipitation estimates for various sample sizes.

reliable, although there would be a chance of some error no matter how many observations are available. The log means and standard deviations for each simulation are presented in Bridges and Haan (1971).

The error probabilities are evaluated by letting Q be a rainfall estimate at probability level x . In general, Q will not equal the true rainfall amount, R . The probability that Q will deviate more than d from R is represented by the shaded area of figure 1 and is the probability of an error of more than d in the rainfall estimate. This probability is determined by selecting some d creating an interval of $R \pm d$ that may or may not contain Q . If we let the probability distribution of rainfall estimates at level x be described by $P(Q)$, the integral of $P(Q)$ from $R-d$ to $R+d$ is the probability that Q is contained in the interval. The probability of an error of d or more is one minus the integrated area. In the case of the log normal distribution, this area can be easily evaluated by taking logarithms of $R-d$ and $R+d$, respectively, and converting them to standard normal values. The area can be evaluated from any standard normal table, and the probability of an error can be calculated. This procedure was carried out for each set of β and m at the 95- and 99-percent probability levels, and the results are presented in tables 2-4, showing the probability of an error greater than d for various sample sizes. The population values shown in table 1 can be used to convert d to a percentage error.

From tables 2-4, one can see that as d decreases for any sample size the probability of an error increases. This is reasonable because the interval in which Q may lie decreases as d gets smaller. It is also shown that as n increases the probability of an error of d or more decreases for a given d . This is due to a decrease in standard deviation as the number of observations increases.

In some cases, the mean of the estimated rainfall amounts did not fall within the interval d . These estimated means were always below the true rainfall amount at a particular probability level because of the procedure used for estimating the parameters of the gamma distribution. For short record length, the parameter \hat{m} was

TABLE 2.—Probability of an error of d or more at 95 and 99 percent for $m=1$ and $\beta=1.00$

n	d (95%)				
	0.1	0.5	1.0	1.5	2.0
10	0.931	0.661	0.365	0.162	0.071
20	.911	.571	.239	.073	.026
30	.880	.447	.126	.030	.009
40	.861	.372	.070	.011	.002
50	.848	.331	.049	.006	.001
100	.783	.161	.006	.000	.000

n	d (99%)				
	0.1	0.5	1.0	1.5	2.0
10	.967	.835	.668	.502	.345
20	.958	.792	.587	.396	.234
30	.946	.732	.485	.282	.141
40	.940	.704	.435	.223	.091
50	.933	.670	.383	.178	.067
100	.919	.609	.299	.114	.037

TABLE 3.—Same as table 2 except $m=2$ and $\beta=1.69$

n	d (95%)				
	0.5	1.0	2.0	3.0	4.0
10	0.827	0.658	0.353	0.137	0.038
20	.756	.529	.191	.044	.010
30	.677	.400	.083	.010	.001
40	.625	.323	.044	.004	.000
50	.587	.271	.025	.001	.000
100	.446	.127	.003	.000	.000

n	d (99%)				
	0.5	1.0	2.0	3.0	4.0
10	.891	.782	.564	.357	.185
20	.834	.672	.382	.169	.054
30	.778	.569	.242	.070	.014
40	.745	.514	.183	.043	.008
50	.726	.479	.145	.025	.003
100	.617	.315	.044	.003	.000

TABLE 4.—Same as table 2 except $m=3$ and $\beta=0.26$

n	d (95%)				
	0.1	0.2	0.3	0.5	1.0
10	0.795	0.600	0.424	0.165	0.008
20	.691	.421	.220	.037	.000
30	.632	.333	.140	.013	.000
40	.582	.267	.093	.006	.000
50	.553	.231	.070	.003	.000
100	.375	.072	.007	.000	.000

n	d (99%)				
	0.1	0.2	0.3	0.5	1.0
10	.826	.727	.597	.364	.051
20	.798	.606	.434	.181	.008
30	.750	.521	.330	.097	.002
40	.714	.460	.264	.059	.001
50	.681	.409	.211	.035	.000
100	.554	.233	.070	.003	.000

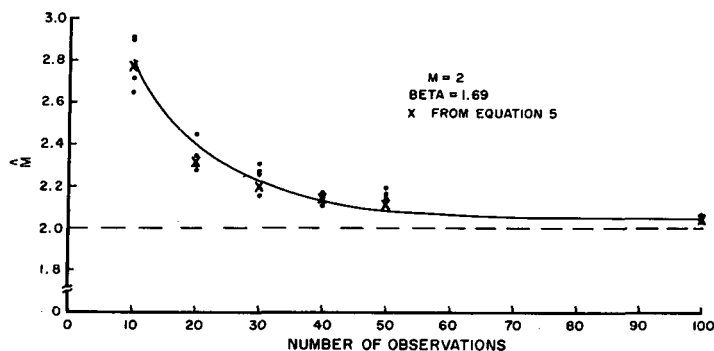


FIGURE 3.—Estimated gamma parameter, \hat{m} , for various sample sizes.

always high and $\hat{\beta}$ was always low. Only when n approached 100 did the estimated parameters ($\hat{\beta}$ and \hat{m}) closely approximate the initial ones. This demonstrates that the procedure for estimating the parameters is biased. Shenton and Bowman (1970) and Bowman and Shenton (1970) discuss the bias in Thom's procedure for estimating the parameters of the gamma distribution. Figures 3 and 4 show the bias in m and β as a function of n for one set of parameters. Bowman and Shenton (1968) present the following approximate relationship for estimating the bias in the parameter m when the method of maximum likelihood is used:

$$E(\hat{m} - m) \approx \frac{3m - \frac{2}{3} + \frac{1}{9}m + \frac{13}{405}m^3}{n - 3} \quad \text{for } n \geq 4, m \geq 1 \quad (5)$$

where $E(\hat{m} - m)$ is the expected bias with an error of less than 1.4 percent. The results of using this relationship for estimating the bias are shown in figure 3 and indicate that the simulated results agree well with the results expected from Bowman and Shenton's relationship. For small values of m (i.e., $m < 0.3$), there is considerable bias in \hat{m} from eq (2) even for samples as large as 150 observations (Bowman and Shenton 1970).

4. DISCUSSION OF RESULTS

Tables 2-4, showing the probability of an error of d or more, present some significant facts to anyone using the gamma distribution in making rainfall estimates. The need for long record lengths in making probability statements can be clearly seen.

As n is increased for any d , the probability of an error is reduced significantly in most cases. It is also noted that very precise values of rainfall estimates corresponding to a small d require extremely long records. As an example, consider the case of $m=1$ and $\beta=1$ at the 95-percent level (table 2). Using the standard procedure for determining the parameter of the gamma distribution, one would have a 36.5-percent chance of missing the rainfall estimate by more than 1 in. if only 10 observations were available. This would be reduced to a 12.6-percent chance

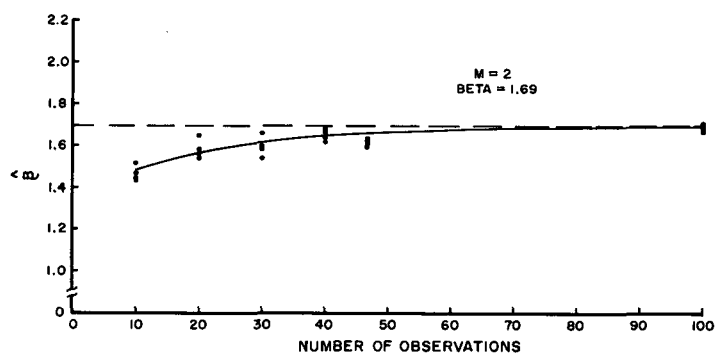


FIGURE 4.—Estimated gamma parameter, $\hat{\beta}$, for various sample sizes.

if 30 observations were available and to a negligible 0.6-percent chance if 100 observations were available. Even with 30 observations, there is a 44.7-percent chance of missing the rainfall estimate by more than 0.5 in. if the true distribution is the gamma distribution with $m=1$ and $\beta=1$.

Error tables such as tables 2-4 can be produced for evaluating the adequacy of any rainfall record. The procedure would be to first estimate the gamma parameters from the available rainfall data. The parameter estimates would then be assumed representative of the population and simulations made from this population. Error probability tables could then be produced. These tables would give the probability of an error of d or more based on the given number of observations or record length if the rainfall distribution truly followed the gamma distribution with the assumed parameters. The probability of a given error could then be determined for the available record length. Since the true gamma parameters would not be known, these probabilities would be approximate but would serve as guide in determining if the available rainfall record was of sufficient length to meet a particular need.

Finally, since the expected bias in the Thom estimators for the parameters of the gamma distribution can be determined, it appears that in future work the parameters should be corrected for bias before probabilities are estimated.

ACKNOWLEDGMENT

The authors acknowledge a review of an earlier draft of this paper by Harold Crutcher, Scientific Advisor, NOAA, Asheville, N.C.

REFERENCES

- Barger, Gerald L., Shaw, Robert H., and Dale, Robert F., "Gamma Distribution Parameters From 2- and 3-Week Precipitation Totals in the North Central Region of the United States," Agricultural and Home Economics Experiment Station, Iowa State University, Ames, 1959, 183 pp.
- Bowman, K. O., and Shenton, L. R., "Properties of Estimators for the Gamma Distribution," Report CTC-1, Union Carbide Corp., Nuclear Division, Oak Ridge, Tenn., 1968, 50 pp.
- Bowman, K. O., and Shenton, L. R., "Small Sample Properties of Estimators for the Gamma Distribution," Report CTC-28, Union Carbide Corp., Nuclear Division, Oak Ridge, Tenn., 1970, 160 pp.

- Bridges, T. C., and Haan, C. T., "Procedure for Determining the Reliability of Precipitation Probabilities Estimated From the Gamma Distribution," *Technical Report 3*, Agricultural Engineering Department, University of Kentucky, Lexington, 1971, 41 pp.
- Friedman, Don G., and Janes, Byron F., "Estimation of Rainfall Probabilities," *Bulletin* No. 332, Agricultural Experiment Station, University of Connecticut, Storrs, Dec. 1957, 22 pp.
- Greenwood, J. Arthur, and Durand, David, "Aids for Fitting the Gamma Distribution by Maximum Likelihood," *Technometrics*, Vol. 2, No. 1, William Byrd Press, Richmond, Va., Feb. 1960, pp. 55-65.
- Hahn, G. J., and Shapiro, S. S., *Statistical Models in Engineering*, John Wiley and Sons, Inc., New York, N.Y., 1967, 355 pp.
- International Business Machines (IBM) Corp., "IBM System/360 Scientific Subroutine Package Version III," *International Business Machines Manual* GH20-0205-4, White Plains, N.Y., 1970, 454 pp.
- Mooley, D. A., and Crutcher, Harold L., "An Application of the Gamma Distribution Function to Indian Rainfall," *ESSA Technical Report EDS 5*, U.S. Department of Commerce, Environmental Data Service, Silver Spring, Md., July 1968, 47 pp.
- Shenton, L. R., and Bowman, K. O., "Remarks on Thom's Estimators for the Gamma Distribution," *Monthly Weather Review*, Vol. 98, No. 2, Feb. 1970, pp. 154-160.
- Strommen, N. D., and Horsfield, J. E., "Monthly Precipitation Probabilities by Climatic Divisions—23 Eastern States," *Miscellaneous Publication* No. 1160, U.S. Departments of Agriculture and Commerce, Washington, D.C., 1969, 141 pp.
- Thom, Herbert C. S., "A Note on the Gamma Distribution," *Monthly Weather Review*, Vol. 86, No. 4, Apr. 1958, pp. 117-122.
- Weisner, C. T., *Hydrometeorology*, Chapman and Hall, Ltd., London, England, 1970, 232 pp.

[Received November 4, 1971; revised April 10, 1972]